RESOURCE USE BY INDIVIDUAL PROJECT - 5/77 THROUGH 4/78

NATIONAL AIM COMMUNITY	CPU (Hours)	CONNECT (Hours)	FILE SPACE (Pages)
1) ACT PROJECT "Acquisition of Cognitive Procedures" John Anderson, Ph.D. Yale University ONR N0014-77-6-0242 (3.5 yrs. 3/77-9/80) 3/78-2/79 $90,000 (*)	153.83	2348.60	2016
2) SECS PROJECT "Simulation & Evaluation of Chemical Synthesis" W. Todd Wipke, Ph.D. U. California, Santa Cruz NIH RR-01059-01 (3 yrs. 7/77-6/80) 7/77-6/78 $94,602 NCI N01-CP-75816 (18 mos.) 3/77-9/78 $58,753 Bayer $5,000 E. Merck $1,500 Sandoz $2,500	360.14	4823.39	7602
3) HIGHER MENTAL FUNCTIONS "Computer Models in Psychiatry and Psychother." Kenneth Colby, M.D. UCLA Biobehavioral Sciences Program funding Proposals pending	48.21	724.20	2374
4) INTERNIST PROJECT "DIALOG: Computer Model of Diagnostic Logic" Jack Myers, M.D. Harry Pople, Ph.D. University of Pittsburgh BHRD MB-00144-04 (4 yrs. 7/74-6/78) 7/77-6/78 $101,000 NIH RR-01101-01 (3 yrs. 7/77-6/80) 7/77-6/78 $160,000	189.68	2735.83	6138

5)	MISL PROJECT "Medical Information Systems Laboratory" Morton Goldberg, M.D. Bruce McCormick, Ph.D. U. Illinois, Chicago Cir. US-PHS-MB00114-04 (4 yrs. 7/74-6/78) 7/77-6/78 $222,487	7.21	251.59	1001
6)	PUFF-VM PROJ (since 10/77) "Pulmonary Function Diag. & Ventilator Management" Edward Feigenbaum, Ph.D. Stanford University John Osborn, M.D. Inst. Medical Sciences, San Francisco NIH approved but unfunded	55.89	1395.85	1234
7)	RUTGERS PROJECT "Computers in Biomedicine" Saul Amarel, Ph.D. NIH RR-00643 (3 yrs. 12/77-11/80) 12/77-11/78 $505,823	32.87	500.10	8437
8)	SCP PROJECT (since 2/78) "Simulation of Comprehension Processes" James Greeno, Ph.D. Alan Lesgold, Ph.D. University of Pittsburgh ONR N0014-78-C-0022 (3 yrs. 10/77-9/80) 10/77-9/78 $62,616 OB-NIE-78-0115 12/77-11/78 $125,900	3.46	61.51	80
9)	AIM PILOT PROJECTS	85.63	1738.97	1911
10)	AIM Administration	16.30	449.84	1155
11)	AIM Users of DENDRAL and MYCIN	27.38	520.42	575
		------	--------	-----
	COMMUNITY TOTALS	980.60	15550.30	32523

STANFORD COMMUNITY	CPU (Hours)	CONNECT (Hours)	FILE SPACE (Pages)
1) AI HANDBOOK PROJECT Edward Feigenbaum, Ph.D. ARPA MDA-903-77-C-0322 (**) (partial support)	100.23	1667.92	1311
2) DENDRAL PROJECT "Resource Related Research Computers and Chemistry" Carl Djerassi, Ph.D. NIH RR-00612 (3 yrs. 5/77-4/80) 5/77-4/78 $218,580	1203.58	20060.60	17460
3) AGE PROJECT (since 9/77) "Generalization of AI Tools" Edward Feigenbaum, Ph.D. ARPA MDA-903-77-C-0322 (**) (partial support)	26.03	775.32	932
4) HYDROID PROJECT "Distributed Processing and Problem Solving" Gio Wiederhold, Ph.D. ARPA MDA-903-77-C-0322 (**)	117.36	2597.67	952
5) MOLGEN PROJECT "Experiment Planning System for Molecular Genetics" Edward Feigenbaum, Ph.D. Joshua Lederberg, Ph.D. NSF MCS76-11649 (2 yrs. 6/76-5/78) 6/77-5/78 $65,610 (*) Nancy Martin, Ph.D. U. New Mexico NSF MCS76-11935 (2 yrs. 7/76-6/78) Total award $68,000 (*)	197.17	3966.60	3125
6) MYCIN PROJECT "Computer-based Consult. in Clin. Therapeutics" Stanley N. Cohen, M.D. Bruce G. Buchanan, Ph.D. NSF MCS77-02712 (2 yrs. 6/77-5/79) 6/77-5/78 $32,357	487.35	6874.76	7892

7) PROTEIN STRUCT MODELING "Heuristic Comp. Applied to Prot. Crystallog." Edward Feigenbaum, Ph.D. NSF MCS-74-23461 (2 yrs. 5/77-4/79) Total award $150,200 (*)	174.21	2933.10	3499
8) PILOT PROJECTS	336.23	6885.61	3728
	-------	--------	-----
COMMUNITY TOTALS	2642.16	45761.58	38899

SUMEX STAFF	CPU (Hours)	CONNECT (Hours)	FILE SPACE (Pages)
1) Staff	661.53	20694.04	14317
2) MAINSAIL Development (since 9/77)	300.63	5125.66	2492
3) Staff affiliates, misc.	33.27	899.74	1219
	-------	---------	------
COMMUNITY TOTALS	995.43	26719.44	18028

SYSTEM OPERATIONS	CPU (Hours)	CONNECT (Hours)	FILE SPACE (Pages)
1) Operations	1986.55	78313.25	75657
	=======	=========	======
RESOURCE TOTALS	6604.74	166344.57	165107

* Award includes indirect costs. All other awards are reported as total direct costs only.

** Supported by a larger ARPA contract MDA-903-77-C-0322 awarded to the Stanford Computer Science Department for the Heuristic Programming Project for the period 8/77-9/79 at a funding level of $765,000 (incl. indirect costs).

2.2.5 NETWORK USAGE

These plots show total terminal connect time for TYMNET and ARPANET users by month since initial connection.

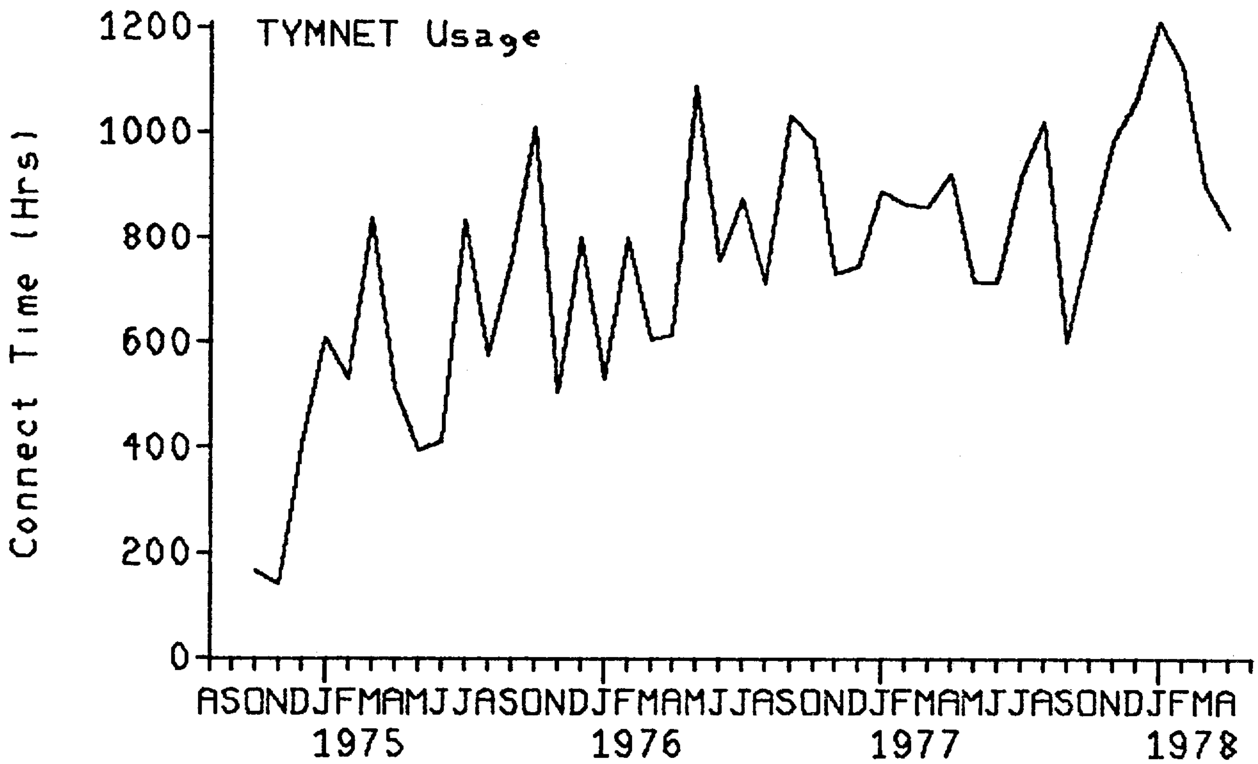

Figure 14. TYMNET Usage Data

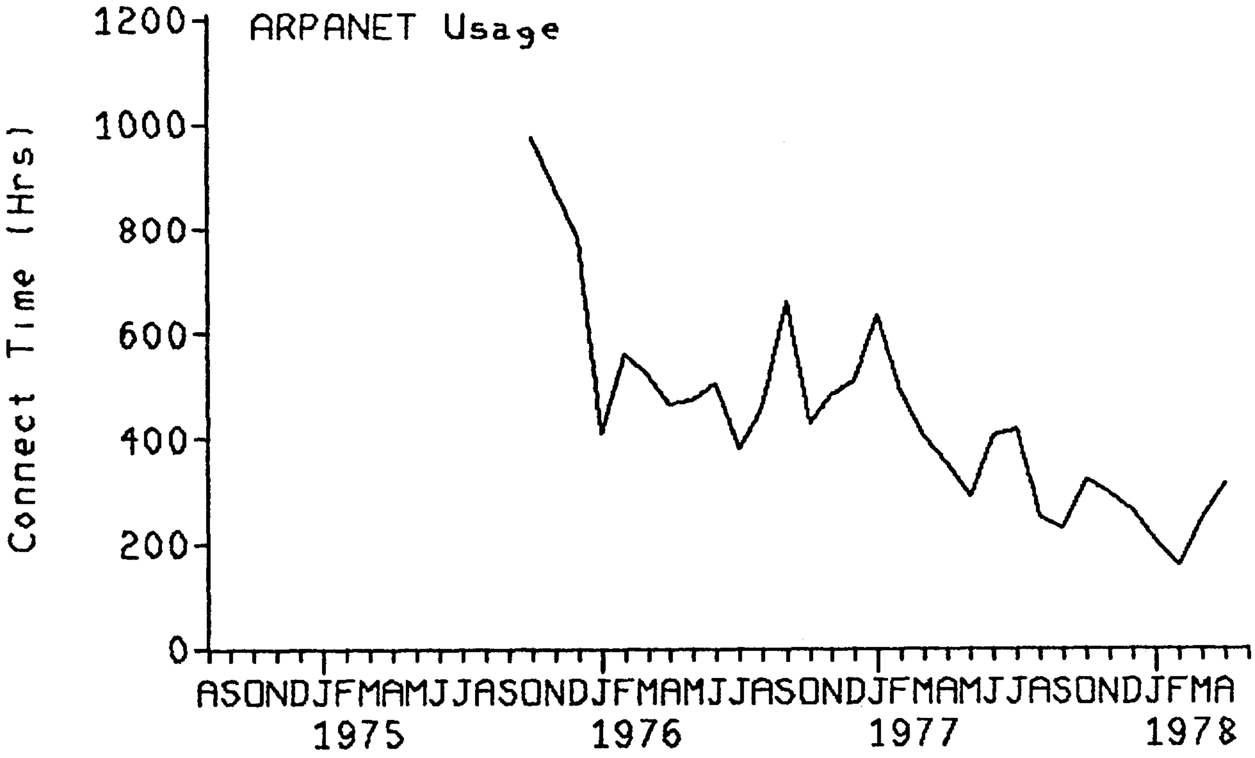

Figure 15. ARPANET Usage Data

2.3 RESOURCE EQUIPMENT SUMMARY

A complete inventory of resource equipment is attached separately as part of the budget material.

2.4 PUBLICATIONS

The following are publications for the SUMEX staff and have included papers describing the SUMEX-AIM resource and on-going research as well as documentation of system and program developments. Publications for individual collaborating projects are detailed in their respective reports (see Section 4 on page 61).

[1] Carhart, R.E., Johnson, S.M., Smith, D.H., Buchanan, B.G., Dromey, R.G., and Lederberg, J, "Networking and a Collaborative Research Community: a Case Study Using the DENDRAL Programs", ACS Symposium Series, Number 19, COMPUTER NETWORKING AND CHEMISTRY, Peter Lykos (Editor), 1975.

[2] Levinthal, E.C., Carhart, R.E., Johnson, S.M., and Lederberg, J., "When Computers Talk to Computers", Industrial Research, November 1975

[3] Wilcox, C. R., "MAINSAIL - A Machine-Independent Programming System," Proceedings of the DEC Users Society, Vol 2, No 4, Spring 1976.

[4] Wilcox, Clark R., "The MAINSAIL Project: Developing Tools for Software Portability," Proceedings, Computer Application in Medical Care, October, 1977, pp. 76-83.

[5] Lederberg, J. L., "Digital Communications and the Conduct of Science - THE NEW LITERACY," Accepted for publication, Proc. IEEE special issue on packet-switched communications.

Mr. Clark Wilcox also chaired the session on "Languages for Portability" at the DECUS DECsystem10 Spring '76 Symposium.

In addition as reported earlier, a substantial effort has gone into developing, upgrading, and extending documentation about the SUMEX-AIM resource, the SUMEX-TENEX system, the many subsystems available to users, and MAINSAIL. These efforts include a number of major documents (such as SOS, PUB, and TENEX-SAIL manuals) as well as a much larger number of document upgrades, user information and introductory notes, an ARPANET Resource Handbook entry, and policy guidelines.

3 RESOURCE FINANCES

3.1 BUDGETARY MATERIALS

The budget for the SUMEX project detailing past actual costs, current year status, and estimates for the next grant year are submitted in a separate document to the NIH.

3.2 RESOURCE FUNDING

The SUMEX-AIM resource is essentially wholly funded by the Biotechnology Resources Program (5). The various collaborator projects which use SUMEX are independently funded with respect to their manpower and operating expenses. They obtain from SUMEX, without charge, access to the computing and, in most cases, communications facilities in exchange for the participation in the scientific and community building goals of SUMEX.

---

(5) Except for participation by Stanford University in accordance with general cost-sharing and for assistance to SUMEX from other projects with overlapping aims and interests.

4 COLLABORATIVE PROJECT REPORTS

The following subsections report on the collaborative use of the SUMEX facility. Descriptions are included for the formally authorized projects within the national AIM and Stanford aliquots and the various "pilot" efforts currently under way. These project descriptions and comments are the result of a solicitation for contributions sent to each of the project Principal Investigators requesting the following information:

I. SUMMARY OF RESEARCH PROGRAM
  A. Technical goals
  B. Medical relevance and collaboration
  C. Progress summary
  D. List of relevant publications
  E. Funding support status (see below for details)

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE
  A. Collaborations and medical use of programs via SUMEX
  B. Sharing and interactions with other SUMEX-AIM projects
     (via workshops, resource facilities, personal contacts, etc.)
  C. Critique of resource management
     (community facilitation and computer services)

III. RESEARCH PLANS (8/78 - 7/81)
  A. Long range project goals and plans
  B. Justification and requirements for continued SUMEX use
     [This section will be of special importance to the Advisory Committee and constitutes your application for continued access.]
  C. Your needs and plans for other computational resources, beyond SUMEX/AIM
  D. Recommendations for future community and resource development

We believe that the reports of the individual projects speak for themselves as rationales for participation; in any case the reports are recorded as submitted and are the responsibility of the indicated project leaders.

4.1 NATIONAL AIM PROJECTS

The following group of projects is formally approved for access to the AIM aliquot of the SUMEX-AIM resource. Their access is based on review by the AIM Advisory Group and approval by the AIM Executive Committee.

4.1.1 ACQUISITION OF COGNITIVE PROCEDURES

ACQUISITION OF COGNITIVE PROCEDURES (ACT)

Dr. John Anderson
Yale University

I. Summary of Research Program

A. Technical goals:

To develop a production system that will serve as an interpreter of the active portion of an associative network. To model a range of cognitive tasks including memory tasks, inferential reasoning, language processing, and problem solving. To develop an induction system capable of acquiring cognitive procedures with a special emphasis on language acquisition.

B. Medical relevance and collaboration:

1. The ACT model is a general model of cognition. It provides a useful model of the development of and performance of the sorts of decision making that occur in medicine.

2. The ACT model also represents basic work in AI. It is in part an attempt to develop a self-organizing intelligent system. As such it is relevant to the goal of development of intelligent artificial aids in medicine.

We have been evolving a collaborative relationship with James Greeno and Allan Lesgold at the University of Pittsburgh. They are applying ACT to modeling the acquisition of reading and problem solving skills. We plan to make ACT a guest system within SUMEX. ACT is currently at the state where it can be shipped to other INTERLISP facilities. We have received a number of inquiries about the ACT system. ACT is a system in a continual state of development but we periodically freeze versions of ACT which we maintain and make available to the national AI community.

C. Progress and accomplishments:

ACT provides a uniform set of theoretical mechanisms to model such aspects of human cognition as memory, inferential processes, language processing, and problem solving. ACT's knowledge base consists of two components, a propositional component and a procedural component. The propositional component is provided by an associative network encoding a set of facts known about the world. This provides the system's semantic memory. The procedural component consists of a set of productions which operate on the associative network. ACT's production system is considerably different than many of the other currently available systems (e.g., Newell's PSG). These differences have been introduced in order to create a system that will operate on an associative network and in order to accurately model certain aspects of human cognition.

A small portion of the semantic network is active at any point in time. Productions can only inspect that portion of the network which is active at the particular time. This restriction to the active portion of the network provides a means to focus the ACT system in a large data base of facts. Activation can spread down network paths from active nodes to activate new nodes and links. To prevent activation from growing continuously there is a dampening process which periodically deactivates all but a select few nodes. The condition of a production specifies that certain features be true of the active portion of the network. The action of a production specifies that certain changes be made to the network. Each production can be conceived of as an independent "demon." Its purpose is to see if the network configuration specified in its condition is satisfied in the active portion of memory. If it is, the production will execute and cause changes to memory. In so doing it can allow or disallow other productions which are looking for their conditions to be satisfied. Both the spread of activation and the selection of productions are parallel processes whose rates are controlled by "strengths" of network links and individual productions. An important aspect of this parallelism is that it is possible for multiple productions to be applied in a cycle. Much of the early work on the ACT system was focused on developing computational devices to reflect the operation of parallel, strength-controlled processes and working out the logic for creating functioning systems in such a computational medium.

We have successfully implemented a number of small-scale systems that model various psychological tasks in the domain of memory, language processing, and inferential reasoning. There was a larger scale project to model the language processing mechanisms of a young child. This includes implementation of a production system to analyze linguistic input, make inferences, ask and answer questions, etc.

The current research is focused on developing mechanisms for the acquisition of skills. In the framework of the ACT system this maps into acquiring new productions and modifying old productions. We have developed learning devices to enable existing productions to create new productions, to adjust the strengths of existing productions, to produce more general variants of existing productions, to produce more discriminant variants of existing productions, and to combine a number of existing productions into a single compact production. We have developed the F version of the ACT system which has these learning facilities. We have so far tested out the system in a number of small learning examples. Current goals involve applying the system to the acquisition of language skills, development of mathematical problem solving skills, and acquisition of initial programming skills.

The basic insight in this research is to model skill acquisition as an interaction between deliberate learning and automatic induction. To the extent that the teacher or the learner is able to understand the skill to be acquired, it is possible for ACT to directly create the necessary productions. However, as a fallback for less structured situations, ACT has automatic induction mechanisms that try to develop the necessary mechanisms by an intelligent trial and error inductive process. Much of our research has gone to identifying the heuristics used by this inductive process. Traditionally, there has been a contrast in psychology between learning with understanding and learning by trial and error. It is now clear to us that most real learning situations involve a mixture and the key to understanding skill acquisition is to understand that mixture.

D. Current list of project publications:

[1] Anderson, J.R. Language, Memory, and Thought. Hillsdale, N.J.: L. Erlbaum, Assoc., 1976.

[2] Kline, P.J. & Anderson, J.P. The ACTE User's Manual, 1976.

[3] Anderson, J.R., Kline, P. & Lewis, C. Language processing by production systems. In P. Carpenter and M. Just (Eds.). Cognitive Processes in Comprehension. L. Erlbaum Assoc., 1977.

[4] Anderson, J.R. Induction of augmented transition networks. Cognitive Science, 1977, 125-157.

[5] Anderson, J.R. & Kline, P. Design of a production system. Paper presented at the Workshop on Pattern-Directed Inference Systems, Hawaii, May 23-27, 1977.

[6] Anderson, J.R. Computer simulation of a language acquisition system: A second report. In D. LaBerge and S.J. Samuels (Eds.). Perception and Comprehension. Hillsdale, N.J.: L. Erlbaum Assoc., 1978.

[7] Anderson, J.R., Kline, P.J., & Beasley, C.M. A theory of the acquisition of cognitive skills. In G.H. Bower (Ed.). Learning and Motivation, Vol. 13. New York: Academic Press, 1979.

[8] Anderson, J.R., Kline, P.J., & Beasley, C.M. Complex Learning. In R. Snow, P.A. Frederico, & W. Montague (Eds.). Aptitude, Learning, an Instruction: Cognitive Processes Analyses. Hillsdale, N.J.: Lawrence Erlbaum Assoc., 1979.

E. Funding:

ONR Contract N0014-77-6-0242 (total period 3/77 - 9/80)
A Model for Procedural Learning
$90,000 3/78 - 2/79 (including indirect costs)

II. Interaction With the SUMEX-AIM Resource

A. & B. Collaborations, interactions, and sharing of programs via SUMEX.

We have received and answered many inquiries about the ACT system over the ARPANET. This involves sending documentations, papers, and copies of programs. We have also used the ARPANET to access and experiment with the production systems at Carnegie Mellon University. The most extensive collaboration has been with Greeno and Lesgold who are also on SUMEX. There is an ongoing effort to help them in their research. Feedback from their work is helping us with system design.

We find the SUMEX-AIM workshops ideal vehicles for updating ourselves on the field and for getting to talk to colleagues about aspects of their work of importance to us.

C. Critique of resource management.

The SUMEX-AIM resource is superbly suited for the needs of our project. We have made the most extensive use of the INTERLISP facilities and the facilities for communication on the ARPANET. We have found the SUMEX personnel extremely helpful both in terms of responding to our immediate emergencies and in providing advice helpful to the long-range progress of the project. Despite the fact that we are on the other side of the continent, we have felt almost no degradation in our ability to do research. We find we can easily list on the terminal a small portion of programs under modification. The willingness of SUMEX to mail listings has also meant we can keep relatively up-to-date records of all programs.

A unique east coast advantage of working with SUMEX is the low loading of the system during the mornings. We have been able to get a great deal of work done during these hours and try to save our computer-intensive work for these hours.

A particularly striking example of the utility of the SUMEX resource was illustrated in the move from Michigan. In the summer of 1976 Anderson moved to Yale and Greeno to Pittsburgh. There was no loss at all associated with having to transfer programs from one system to another. At Yale we were programming the day after we arrived. The SUMEX link has also permitted continued collaboration with Greeno. We are planning a permanent move to Carnegie-Mellon this summer and happily anticipate it will be as painless.

III. Research Plans (8/78-7/81)

A. Long-range user project goals and plans:

Our long-range goals are: (1) Continued development of the ACT system; (2) Application of the system to modeling of various cognitive processes; (3) Dissemination of the ACT system to the national AI community.

1. System Development We have completed the F version of the ACT system. We are currently applying or intend to apply the ACT system to modeling the acquisition and/or performance of cognitive skills in the areas of language comprehension and generation, inferential reasoning, reading skills, mathematical problem solving, and computer programming. It is hard to anticipate now all the impact of these explorations for design decisions in later versions of ACT. However, it is clear even now that a number of developments are needed. We want to make ACT more appropriate as a language for programming cognitive skills. This involves such things as development of more powerful control conventions, simplification of syntax, and introduction of direct programming features (such as comparison of quantity magnitudes) that can only be obtained indirectly in ACTF. We also want to introduce more efficient implementation techniques to replace some of the simple devices that were used to enable us to rapidly complete the system. These rearchitecture efforts have to be done within the

constraints of psychological plausibility, but we have a theoretical commitment to the conjecture that good implementation design is predictive of good psychological mechanisms. We are currently implementing a new system--G version--which will incorporate these ideas and any additional insights that will come out of our experimentation with ACTF.

2. Application to Modeling Cognitive Processes. We anticipate a gradual decrease in the amount of effort that will go into system development and an increase in the amount of effort that will go into application of the system for modeling. We mentioned above the modeling efforts that we are using to assess the suitability of the ACTF system. We have long-range commitments to apply the ACT learning model to the following three topics: Acquisition of language (both first and second language acquisition); acquisition of programming skills; acquisition of problem solving skills in the domain of geometry. We find each of these topics to be considerable interest in and of themselves, but they also will serve as strong tests of the learning model. We are hopeful that the systems that are acquired by ACT will satisfy computational standards of good artificial intelligence. Therefore, in future years we would also be interested in applying the ACT model to acquisition of cognitive skills in medically related domains such as diagnosis or scientific inference. SUMEX would be an ideal location for collaboration on such a project.

3. Dissemination of the ACT Project We have a commitment to making the ACT system available to anyone in the national community who has access to the necessary computer resources. This is partially to provide a service in that ACT is a medium for psychological modeling. However, it is also self-serving in that the use of other people make of ACT has important feedback in assessing design decisions. In light of limitations on the SUMEX resource, we have decided not to allow extensive use of ACT by other researchers through our SUMEX account. We feel that extensive use of the ACT system in SUMEX by another researcher must have the status of an independent project and must be able to justify independently its use of the SUMEX-AIM resource. The current system being supported for use by other researchers is ACTE but we are in the process of updating our supported system to ACTF.

B. Justification for continued use of SUMEX:

We feel that the justification for our use of SUMEX has only been strengthened since the time of our original application for user status. The project meets a number of criteria for SUMEX relevance: The project is concerned with cognitive modeling which is a SUMEX goal. The project is also developing an AI tool which can be used to help automate various medically-relevant tasks. We also think we are the type of need that the SUMEX facility was designed to meet. That is, we do not have nearly as powerful computing facilities local at Yale; we are non-local user; we are using SUMEX as a base for collaborating with scientists in other parts of the country; and we are trying to develop a system that will be of general use.

Our future move to Carnegie Mellon raises some interesting issues about the SUMEX resources. The availability of the SUMEX resource makes the move easy and allows the research to go full steam ahead. The fact that Carnegie-Mellon is on the ARPANET will reduce our cost (no TYMNET charges) and allows us to get immediate listings which had been the one deficit of being a distant user. On the other hand, the greater Carnegie resources may diminish the need for SUMEX.

We have not carefully explored the quality of the Carnegie system versus SUMEX but an important obstacle for us is the lack of INTERLISP at Carnegie. Carnegie will have INTERLISP available within the next two years. At such time it may be appropriate to enter discussion with SUMEX about how to balance our lesser needs for SUMEX with our cost to SUMEX and both of these factors with the role we play in the SUMEX community.

C. Comments and suggestions for future resource goals:

We would, of course, be delighted if the computational capacity of the SUMEX facility could be increased. The slowness of the system at peak hours is a limiting factor although it is not grievous. This problem is perhaps less grievous for us than Stanford-based users because of our ability to use morning hours. We do not feel any urgent need for development of new software.

4.1.2 CHEMICAL SYNTHESIS PROJECT

SECS - Simulation and Evaluation of Chemical Synthesis

Principal Investigator: W. Todd Wipke
Board of Studies in Chemistry
University of California at Santa Cruz

Coworkers: S. Krishnan, C. Buse, and M. Huber (Postdoctoral Fellows)
G. Ouchi and D. Dolata (Grad students)

I. SUMMARY OF RESEARCH PROGRAM

A. Technical Goals

The long range goal of this project is to develop the logical principles of molecular construction and to use these in developing practical computer programs to assist investigators in designing stereospecific syntheses of complex bio-organic molecules. Our specific goals this past year focused on strategic control of the SECS program, on implementing strategies based on symmetry and potential symmetry, on developing ways to treat the steric factors involved in acyclic reaction centers, on converting SECS from F40 FORTRAN to F10 FORTRAN (the current supported FORTRAN), and on demonstrating that computer synthesis techniques are also applicable to metabolism. In addition, we wanted to add a library lookup capability, and to improve the user interaction with the SECS program.

B. Medical Relevance and Collaboration.

The development of new drugs and the study of how drug structure is related to biological activity depends upon the chemist's ability to synthesize new molecules as well as his ability to modify existing structures, e.g., incorporating isotopic labels or other substituents into biomolecular substrates. The Simulation and Evaluation of Chemical Synthesis (SECS) project aims at assisting the synthetic chemist in designing stereospecific syntheses of biologically important molecules. The advantages of this computer approach over normal manual approaches are many: 1) greater speed in designing a synthesis; 2) freedom from bias of past experience and past solutions; 3) thorough consideration of all possible syntheses using a more extensive library of chemical reactions than any individual person can remember; 4) greater capability of the computer to deal with the many structures which result; and 6) capability of computer to see molecules in graph theoretical sense, free from bias of 2-D projection.

The objective of using SECS in metabolism is to predict the plausible metabolites of a given xenobiotic in order that they may be analyzed for possible carcinogenicity. Metabolism research may also find this useful in the identification of metabolites in that it suggests what to look for. Finally, it seems there may even be application of this technique in problem domains where one wishes to alter molecules so certain types of metabolism will be blocked.

C. Progress and Accomplishments

RESEARCH ENVIRONMENT: At the University of California, Santa Cruz, we have a GT40 graphics terminal connected to the SUMEX-AIM resource by a 1200 baud leased line (the leased line supported by SUMEX). On 1 May 1978 a new GT46 graphics terminal was installed in addition. We also have a TI725, a TI745, a CDI-1030, a DIABLO 1620, and an ADM-3A terminal which were used over the University tie-line system to SUMEX, but now are awaiting leased lines since the University of California did away with the tie-line system. UCSC has only a small IBM 370/145, a PDP-11/45 and 11/70 (the latter are limited to small student time-sharing jobs of 12 K words per user), all of which are unsuitable for this research. We hope through the GT46 system, to be able to transfer files from SUMEX to the UCSC PDP-11 for local printing, and possibly local magnetic tape handling (currently all input and output except graphical must be done at Stanford.) The SECS laboratory is located in the same building as the synthetic chemists at Santa Cruz so there is very facile interaction.

STRATEGIC CONTROL: We feel, and feedback from users supports this, that there is a great need for the user of SECS to control and direct the synthesis planning, if the user so desires. For the purpose of this discussion, a strategy is a general principle which helps guide one in generating a simple synthesis. Strategies are based on symmetry, mathematical considerations of yield, economy of operations, etc. When a strategy is applied to a particular synthetic target molecule, it generates goals. Goals are described only in terms of molecular structural changes or features, and may not, for example, refer to reactions. Thus, strategies create goals, and both are completely independent of the reaction library.

We had previously created a list-structured language for describing goals to allow manual introduction of goals which can then direct selection of relevant transforms to only those that satisfy the specified goals. We have continued to improve the human interface to that module, but the majority of our current work is on developing modules which implement various high level strategies and thus automatically create goals. The chemist of course will still have the opportunity to modify these automatically created goals. The reason for wanting to create them automatically is that the chemist would never be able to consider all possible strategies and much of the creativity in a synthesis is in selection of the basic strategy and resulting goals. And the reason for having goals is that it gives SECS a sense of purpose and justifies therefore devoting more resources to following those goals and trying to achieve them, sometimes to the exclusion of other chemical transformations. The power of the goal list has led to some unexpected capabilities. For example, by specifying a certain set of changes in the target molecule and that only those transforms which satisfy that goal list are allowed to be applied, one can find out whether SECS has a reaction in its library capable of making those changes. The following paragraphs describe some of the current strategy work.

STRATEGIES BASED ON SYMMETRY: Based on analyses of many literature syntheses, we have found several key strategies related to symmetry. One of these involves trying to break the structure into two or more identical fragments. The advantages of a synthesis utilizing identical or similar fragments result both from a minimization of the number of synthetic steps, and from the principle of convergent synthesis (if the identical fragments constitute a major part of the target compound). This analysis takes place on the strategic level, meaning that SECS considers the non-redundant disconnections which generate fragments, regardless of whether any applicable reactions exist which could actually form the bond broken in the analysis. Identical fragments are recognized by comparing their SEMA canonical names. In order to find similar fragments we first had to define what "similar" meant and then to define a function to calculate a metric on that space. The current function considers atom types, bond types, and stereochemistry in addition to connectivity, and it includes weighting factors for each term. At the end of the analysis, fragment sets containing identical or similar fragments are shown to the user and can then be converted into GOALS. Later, SECS uses those GOALS to find transforms that can achieve the desired constructions.

When molecules that have been synthesized from identical or similar fragments (literature) are analyzed by the program, it so far has always been able to detect the proper fragmentations. Interestingly, in analyzing natural products, SECS "discovered" the isoprene rule since nature uses this strategy too! This module generates some very challenging suggestions, even though there may not be known reactions to implement the suggestions. In fact, this now provides SECS with a rationale which in the future might suggest a need for a new reaction.

SUBGOALS: When a chemical transform has a high priority and seems to be able to satisfy a goal on the goal list the transform is "relevant", but still may not be "applicable" owing to some mismatch between what the transform requires and what the operand structure has. This mismatch can spawn a SUBGOAL to change the structure until this transform is applicable. The first utilization of subgoals in SECS is for automatic functional group interchange (FGI). Mismatches in the identity of functional groups are easier to correct than mismatches in the carbon skeleton. The program now recognizes these functional group mismatches, generates subgoals, tries to satisfy the subgoals, and then goes back to the original goal. This is equivalent to a "look-ahead" and it may lead to sequences of several reactions, but SECS invents the sequences, they are not preprogrammed. At the time of this writing, this module works well on small molecules, but on complex ones, the number of subgoals created is too large. Methods of combating this problem are under consideration.

STRUCTURE DICTIONARY: "Is this compound commercially available?", "Is this a new compound?", and "Is this a known carcinogen?" are some of the questions a synthetic chemist asks in designing a synthesis. In order to be able to answer questions like these rapidly regardless of how large the chemical dictionary may be, we developed an efficient search method using a hash of the SEMA canonical name. We studied the efficiency of several hash functions, with various sized tables and varying amounts of detail in the SEMA name. On a typical file of compounds from a synthesis, our hash function randomized the keys as well as a random number generator. SECS uses this technique now in finding if a precursor already exists in the synthetic tree. We have also used this for finding common biosynthetic intermediates by checking the precursor against other "trees" which were generated in previous analysis of other structures. We plan to add other libraries of compounds in the future.

STEREOCHEMICAL INDUCTION: In the synthesis of macrocyclic natural products such as erythronolide A, or maytansine, stereochemical control is of prime importance. This can be achieved by using small rings, or by using stereospecific acyclic carbonyl addition reactions which follow Cram's rule. We have implemented a module to evaluate Cram's rule and alter the reaction's priority value accordingly. The module has been tested using over 50 reactions from the literature.

EXPLANATION CAPABILITY: We feel that in order to gain the confidence of the user, SECS should be able to explain the source of its knowledge and some of its reasoning. Our first step in this direction is a rapid retrieval of literature references for transforms. The second step is modeled after the question-answering of MYCIN. For some time now, SECS has recorded the types of questions chemists ask. A parser for those questions is now underway with the answer generator to follow. We hope to be able to answer questions such as "Why did you not use the Wittig reaction here?" or "Why is the priority of this reaction so high?"

METABOLISM PREDICTION: Numerous structurally different chemical compounds have been found to induce neoplasia in man and animals. In many cases these chemical carcinogens are metabolically activated by mammalian enzyme systems to their ultimate reactive and toxic structure. Many of the mechanisms involved in this "bioactivation" process are known or are in the process of being discovered. Thus, it is now possible based on the structure of a compound and a through knowledge of biotransformations to make rational predictions of the plausible metabolites of a compounds produced in a mammalian system. To study the metabolic activation of compounds we are creating a computer assistant which will generate the plausible metabolites of a compound utilizing the biotransformations known to occur in mammalian systems.

A new computer program called XENO for the metabolism of xenobiotic compounds has been developed based on technology from computer synthesis project. However, since metabolism is being simulated in the forward direction, whereas organic synthesis is simulated in the reverse direction, the XENO program is quite different in logic from SECS, although both use ALCHEM as a representation for reactions. The XENO data base of biotransforms was developed by careful survey of metabolism literature and consultation with a committee of metabolism experts at NIH. We selected a mechanistic representation of metabolic processes which means a small data base suffices to represent most of the known processes. A critical evaluation of XENO by a panel of experts in Bethesda, Md. in February 1978 concluded that the data base of biotransforms must be considerably expanded, but even now it is able to raise some interesting questions of alternative metabolic pathways, etc. XENO is currently running on SUMEX-AIM. Shown below is part of the analysis XENO performed on benzo(a)pyrene.

```
:TREE
  1:  2   3   6   9   12  13  14  15  16  17  19  24  26
      27  28  29  30  31  32  33  34  35  36  37  38  39
  2:  40  41  44  46  47  48  52  53
 40:  54  55  56  57  58  59  61  66  68  69  70  71  72
      73  74  75  76  77  78  79  80  81  82  83  85  86
      87  88  89  90  91  92  93  94  95  96
```

1

```
      C=C
     /   \
  C-C     C=C
 =   =   /   \
C     C-C     C-C
 \   /   =   =   =
  C=C     C-C     C
     \   /   \   /
      C=C     C=C
```

Benzo(a)pyrene

2

```
      C=C
     /   \
  C-C     C=C     O
 =   =   /   \   /|
C     C-C     C-C |
 \   /   =   =   \|
  C=C     C-C     C
     \   /   \   /
      C=C     C=C
```

Priority = 120

40

```
          C-C
         =   =
      C-C     C
     =   \   /
    C     C=C
     \   /   \
      C=C     C
     /   \   =
  C-C     C-C
 =   =   =
C     C-C
 \   /
  C-C
 /   \
O     O
```

Priority = 105

54

```
          C-C
         =   =
      C-C     C
     =   \   /
    C     C=C
     \   /   \
O     C=C     C
|\   /   \   =
| C-C     C-C
|/   =   =
C     C-C
 \   /
  C-C
 /   \
O     O
```

Priority = 240 *

55

```
          C-C
         =   =
      C-C     C
     =   \   /
    C     C=C
     \   /   \
O     C=C     C
|\   /   \   =
| C-C     C-C
|/   =   =
C     C-C
 \   /
  C-C
 /   \
O     O
```

Priority = 140 *

D. List of Current Project Publications

S.A. Godleski, P.v.R. Schleyer, E. Osawa, and W.T. Wipke, "The Systematic Prediction of the Most Stable Neutral Hydrocarbon Isomer," J. Am. Chem. Soc., 99, 0000(1978).

W. T. Wipke, G. Smith, F. Choplin, and W. Sieber, "SECS--Simulation and Evaluation of Chemical Synthesis: Strategy and Planning," ACS Symposium Series No. 61, 97-127 (1977).

W.T. Wipke, "Computer Planning of Research in Organic Chemistry," in Computers in Chemical Education and Research, ed. E.V. Ludena, N.H. Sabelli, and A.C. Wahl, Plenum Press, N.Y., 1977, pp. 381-391.

W.T. Wipke, S. Krishnan, and G.I. Ouchi, "Hash Functions for Rapid Storage and Retrieval of Chemical Structures," J. Chem. Info. and Computer Sci., 18, 32 (1978).

F. Choplin, R. Marc, G. Kaufmann, and W.T. Wipke, "Computer Design of Synthesis in Phosphorus Chemistry. Automatic Treatment of Stereochemistry," J. Chem. Info. and Computer Sci., 18, 000 (1978).

F. Choplin, R. Dorschner, G. Kaufmann, and W. T. Wipke, "Computer Graphics Determination and Display of Stereoisomers in Coordination Compounds," in press.

F. Choplin, C. Laurenco, R. Marc, G. Kaufmann, and W.T. Wipke, "Synthese Assistee par Ordinateur en Chimie des Composes Organophosphores," Nouveau J. de Chimie, in press.

W.T. Wipke, G. Ouchi, and S.Krishnan, "Simulation and Evaluation of Chemical Synthesis - SECS. An Application of Artificial Intelligence Techniques," Artificial Intelligence, in press.

M. Spann, K. Chu, W.T. Wipke, and G. Ouchi, "Computer-Aided Prediction of Metabolites," in press.

E. Funding Status

1. Resource-Related Research: Biomolecular Synthesis
   PI: W. Todd Wipke, Associate Professor, UCSC
   Agency: NIH, Research Resources
   No: RR01059-01
   7/1/77-6/30/80 $227,816 TDC
   7/1/77-6/30/78 $94,602 TDC

2. Computer-Aided Prediction of Metabolites for Carcinogenicity Studies
   PI: W. Todd Wipke
   Agency: NIH, National Cancer Institute
   No: N01-cp-75816
   3/8/77-9/7/78 $44,146 TDC
   3/8/78-9/7/78 $14,607 TDC

3. Computer Synthesis, Unrestricted
   PI: W. Todd Wipke, Associate Professor, UCSC
   Agency: Sandoz, Ltd. $2500
   Agency: Bayer $5000
   Agency: E. Merck $1500

II. INTERACTIONS WITH SUMEX-AIM RESOURCE

A. Collaborations and Medical Use of Programs via SUMEX

SECS is available in the GUEST area of SUMEX for casual users, and in the SECS DEMO area for serious collaborators who plan to use a significant amount of time and need to save the synthesis tree generated. Much of the access by others has been through the terminal equipment at Santa Cruz because graphic terminals make it so much more convenient for structure input and output. We have assisted Professor J.E. McMurry of UCSC in his synthetic work towards aphidicoline and digitoxigenin (Total Synthesis of Cardiac Aglycones, HL-18118) using the model builder of SECS for evaluating plausible modes of ring closure. Numerous visitors to UC Santa Cruz have tried their own problems on the SECS program, generally taking away at least a couple of new ideas for research. Professor Ken Williamson of Mt. Holyoke College has made arrangements to access SECS to obtain structures for C-13 nmr analysis, and a student at the University of Mass. Amherst has made arrangements to do several analyses on SECS as an independent research project, the results to be tested in the laboratory. The entire collaboration between Drs. Ted Gram of Guarino's lab, Lance Pohl from Gillette's lab, Dhiren Thakken and Harukiko Hagi from Jerina's lab, Ken Chu and Sid Siegel (chemical carcinogenesis), and Mel Spann (National Library of Medicine) in Bethesda would not be possible without access to XENO through SUMEX.

Synthetic chemists are beginning to come to us for a SECS analysis before beginning a laboratory synthesis. Dr. McMurry for example did a rather complete analysis of morphine before launching his recently successful synthesis. We have also collaborated in the biogenesis work with Professor Phil Crews (UCSC) in marine natural product biogenesis. Dr. Wipke has also used several SUMEX programs such as CONGEN in his course on Computers and Information Processing in Chemistry.

B. Examples of Sharing, Contacts and Cross-fertilization with other SUMEX-AIM projects

Dr. Wipke spent the Winter Quarter on sabbatical at Stanford and regularly attended the SIGLUNCH seminars of the Heuristic Programming Project. We have had several discussions with the MYCIN group about our interest in an explanation capability for SECS. The AIM conference at Rutgers each year has been extremely valuable in generating ideas of new ways to apply current developments in AI to the problem of organic synthesis. Finally, it is impossible to count the daily exchanges that occur between researchers in the SECS group and other members of the AIM community on things related to languages, conferences, papers, seminars, and program sharing. Now that our GT46 is installed, for example, we have been communicating with Achenbach at Stanford regarding the AMOK file transfer system which will help us get local printing of files.